

**MULTI KOLINIERITAS DALAM REGRESI MULTIPLE LOGISTIK**

**Hery Tri Sutanto**  
**Jurusan Matematika MIPA Unesa**  
**Surabaya**

**Abstrak**

Adanya korelasi yang tinggi antar variabel bebas menandakan adanya kolinearity dalam model regresi logistik. Untuk ukuran sampel yang kecil akan melemahkan signifikansi dari masing-masing variabel bebas padahal model secara keseluruhan kuat signifikansinya. Multikolinearity menghasilkan interpretasi yang salah tentang estimator koefisien regresi logistik. Hal ini berakibat kesimpulan tentang hubungan antara variabel bebas dengan variabel tak bebas salah.

Kata Kunci: Kollinearity , regresi logistik, parameter, variabel bebas dan variabel tak bebas

**1. Pendahuluan**

Meskipun multikollinieritas dalam regresi multipel logistik tidak seluas multikollinieritas dalam regresi linear berganda. Dilain pihak hanya sedikit referensi yang membicarakan multikollinieritas dalam regresi multiple logistik. Pada multikollinieritas dalam regresi linear berganda dibicarakan bahwa jika ada korrelasi tinggi dengan  $y$  akan menyebabkan nilai  $t$  kecil demikian pula pada multikollinieritas dalam regresi multiple logistik terdapat korelasi tinggi dan nilai  $t$  kecil.

Kita harus memilih satu atau lebih indikator dari multikollinieritas dalam regresi logistik. Jika variabel-variabel bebasnya semua kontinyu maka ada korrelasi Pearson dan VIF (Varians Inflation Factors) bisa digunakan. Masalah yang timbul akan lebih kompleks jika semua variabel bebasnya tidak kontinue. Hosmer dan Lemeshow (1989 hal 132) membicarakan secara singkat tentang cara mendeteksi kollinieritas dalam regresi logistik yang mirip dalam regresi linear berganda. Dapat dilihat dalam makalah ini bahwa kollinearitas ada dan standard error yang tinggi.

## Contoh1

Perhatikan data dalam table 1 dengan dua variabel bebas biner. Ada dua cara untuk menduga parameter (koefisien) setiap variabel bebas, yaitu metode maksimum likelihood dan ada estimasi pendekatan tepat dalam tabel2. Perhatikan perbedaan yang besar antara nilai pendekatan dan p exact yang menggambarkan perbedaan ini yang menyatakan sampel ukuran  $n=24$  kecil. Nilai p exact

Yang menyatakan bahwa  $X_1$  dan  $X_2$  tidak diperlukan dalam model. Kalau kita amati tabel1 kita lihat bahwa nilai  $X_1$  dan  $Y$  yang berkaitan ada 18 dari 24 kasus dan nilai  $X_2$  dan  $Y$  yang berkaitan ada 18 dari 24 kasus yang ada. Jadi variabel  $X_1$  dan  $X_2$  ada hubungan yang kuat. Kita dapat melihat  $X_1$  dan  $X_2$  sepakat bahwa 16 dari 24 kasus sehingga  $X_1$  dan  $X_2$  terasa ada hubungan. Hal itu akan membantu mengecek satu formal atau lebih multicollinearity. Hosmer dan Lemeshow (1989, hal 131) menyatakan bahwa beberapa paket software untuk mengecek multicollinearity. Salah satu kemungkinan dua variabel bebas binary yang akan digunakan Cohen's kappa yang asseses the extent of agreement antara dua variabel binary relative apa yang akan yang diharapkan due of chance. Nilai statistik ini menggunakan  $X_1$  dan  $X_2$  adalah 0,338.

Tabel 1 Data untuk mengilustrasikan multikolinearitas dalam Regresi multiple logistik

$X_1$	$X_2$	Y
1	1	1
1	1	1
1	0	1
0	1	1
1	1	1
0	0	0
0	0	0
1	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	1	0
0	0	0
1	1	1
1	0	1
1	1	0
0	1	1
1	1	1
1	1	1

1	1	1
0	0	1
1	0	0
1	0	1

## Contoh 2

Terjadinya multikolinearitas jika ada dua variabel independent lebih dalam model ditentukan oleh kombinasi linear dari variabel independent lain dalam model. Misal kita mempunyai suatu masalah multikolinearitas jika kita telah mengukur tinggi sesuatu dalam inchi dan mengukur tinggi sesuatu dalam feet dalam model yang sama. Derajat multikolinearitas dapat bervariasi dan mempunyai effect yang berbeda dalam model. Jika terjadi multikolinearity secara lengkap yaitu jika salah satu variabel independent merupakan suatu kombinasi linear lengkap dari variabel independent lain, hal yang demikian tidak mungkin diperoleh estimator koefisien regresi secara tunggal dengan semua variabel independent dalam model. Mengapa stata dilakukan dalam kasus ini adalah dengan cara membuang suatu variabel yaitu kombinasi linear lengkap dari variabel lain, meninggalkan hanya variabel-variabel yang tidak exactly kombinasi linear dari variabel lain dalam model yang diasumsikan estimator koefisien regresi tunggal. Misal kita dapat membuat artificial yang merupakan suatu variabel baru yang dinamakan parli sebagai jumlah dari yr-rnd dan meals. Perhatikan hanya bermaksud dari contoh ini dan membuat variabel-variabel perli yang menunjukkan mengapa stata jika terjadi multicollinearity lengkap.

Multicollinearity biasanya terjadi sejak ada korrelasi diantara variabel independent yang merupakan indikator dari collinearity. Jika terjadi beberapa multikolinearity secara terpisah, maka standard error untuk koefisien regresi cenderung sangat besar dan kadang-kadang estimasi koefisien regresi logistik yang unreliabelnnya tinggi. Mari perhatikan contoh berikut. Dalam model ini variabel dependent akan hiqual dan variabel independent terdiri dari avg-ed, yr-rnd, meals, full dan interaksi antara yr-rnd dan full, yxfull.

Untuk mendeteksi multikolinearity dengan menggunakan tolerance merupakan indikator berapa banyaknya collinearity dan VIF (Variance Inflation Faktorian) merupakan indikator besarnya inflation dari standard error yang disebabkan oleh collinearity. Tolerance untuk suatu variabel adalah  $1 - R^2$ . Sedangkan besarnya VIF sama dengan  $\frac{1}{\text{tolerance}}$ .

Collin avg-ed yr-rnd meals full yxfull

#### Collinearity Diagnostics

Variabel	Sort			Cond	
	VIF	VIF	Tolerance	Eigenval	Index
Avg-ed	3,28	1,81	0,3050	2.7056	1.0000
Yr-rnd	35,53	5,96	0,00281	1.4668	1.3581
Meals	3,80	1,95	0,2629	0,6579	2,0279
Full	1,72	1,31	0,5819	0,1554	4,1720
Yxfull	34,34	5,86	0,0291	0,0144	13.7284
Mean VIF	15.73		condition Number 13.7284		

Dari output diatas diperoleh nilai tolerance untuk variabel yxfull sebesar 0,0291 dan nilai VIF sebesar 34,34. Kita dapat menghasilkan regresi berikut ini:

Regrese yxfull full meals yr-rnd avg-ed

Source	SS	df	MS	Number of obs = 1158
Model	1128915.43	4	282228.856	F(4,1153)= 9609,80
Residual	33862,2808	1153	29.3688472	R-squared= 0,9709
Total	1162777,71	1157	1004.9937	Adjs R-squared=0,9709
				Root MSE = 5,4193

Yxfull	Coef	Std Error	t	P> t	(95 % Conf. Interval)	
Full	0,2313279	0,0140312	16.49	0,000	0,203983	0,2588574
Meals	0,00088	0,099863	0,09	0,930	0,0204733	0,0187134
Yr-rnd	83.10644	0,4408941	188,50	0,000	82,2414	83,97149
Avg-ed	0,4611434	0,3744277	1,23	0,218	1,195779	0,2734925
Cons	19,38205	2,100101	9,23	0,000	23,5025	15,2616

Berdasarkan output diatas  $R^2=0,9709$ . Karena tolerance  $1-0,9709=0,0291$  maka nilai

$$VIF = \frac{1}{0,0291} = 34,36$$

#### Pustaka

1. Ryan, Thomas, 1997. Modern Regression Methods, John Wiley & Sons, New York.